*Research Article*
# Dual Divergence Estimators of the Tail Index

## Salim Bouzebda[1] and Mohamed Cherfi[2]

[1] *Laboratoire de Mathématiques Appliquées, Université de Technologie de Compiègne,*
  *BP 529, 60205 Compiègne Cedex, France*
[2] *Department of Mathematics, Faculty of Sciences, Hassiba Benbouali University of Chlef,*
  *BP 151, 02000 Chlef, Algeria*

Correspondence should be addressed to Mohamed Cherfi, mohamed.cherfi@gmail.com

The main purpose of the present paper is to propose a new estimator of the tail index using $\phi$-divergences and the duality technique. These estimators are explored with respect to robustness through the influence function approach. The empirical performances of the proposed estimators are illustrated by simulation.

## 1. Introduction

In extreme value statistics emphasis lies on the modelling of rare events, mostly events with a low frequency but a high impact. Common practice is to characterize the size and frequency of such extreme events mainly by the extreme value index $\gamma$, and here the main problem is to estimate the unknown parameter $\gamma$. Since only the upper tail of the distribution is involved, it is reasonable to construct estimators of $\gamma$ based on the top extreme values of a sample $X_1, \ldots, X_n$. The most commonly used estimator of the kind is that proposed by Hill [1]. We mention that the most prominent estimators of this real-valued parameter $\gamma$ are maximum likelihood estimators of specific parametric models which are fitted to excesses over large thresholds (see [2]). Indeed, alternatives to the Hill estimator are discussed by Smith [2]. One of his conclusions (see, e.g., [2], pp. 1181-1182) is that in general the Hill estimator compares favourably with other competitors. In general, these maximum likelihood estimators often prove to be highly efficient, though nonrobust against deviations of the actual distribution from the assumed parametric model. This is, for instance, the case in the presence of outliers or suspicious data, where the performance of the maximum likelihood estimators and the quality of the corresponding estimates of the tail index are often seriously affected.

It is known that the maximum likelihood estimation is very sensitive to deviations from theoretical distributions which is, not surprisingly, the case for the class of heavy-tailed distributions, and fails to provide a reasonable parameter estimation; refer to Alexander [3] among others.

Robustness is an important issue in extreme value theory; see for instance DellAquila and Embrechts [4]. As shown in Brazauskas and Serfling [5], small errors in the estimation of the tail index can already produce large errors in the estimation of quantities based on the tail index $\gamma$. Hence, to overcome the lack of robustness to outliers of this kind of estimators, some robust methods for extreme values have already been discussed in recent literature. The interested reader may refer to Brazauskas and Serfling [6] for robust estimation in the context of strict Pareto distributions. Dupuis and Field [7], respectively, Peng and Welsh [8] and Juárez and Schucany [9], derived robust estimation methods for the case where the observations follow a generalized extreme value distribution, respectively, a generalized Pareto distribution, light or heavy tailed. Vandewalle et al. [10] considered a robust estimation method based on the minimization of the integrated squared error criterion using an incomplete density mixture model; Kim and Lee [11] used the minimum density power divergence approach of Basu et al. [12] to estimate the tail index in the dependent case; more recently Hubert et al. [13] proposed a method to detect outliers that can influence the Hill [1] estimator.

In this paper, we propose a new robust tail index estimation procedure, based on $\phi$-divergences and the duality technique, for the semi-parametric setting of Pareto-type (or heavy-tailed) distributions. So here, the strict Pareto distribution is assumed to hold only asymptotically, that is, for excess distributions over high enough threshold values. The proposed method extends the maximum likelihood procedure, and, it will be seen that the last method corresponds to the particular choice of the $KL_m$-divergence which leads to the Hill [1]'s etimator.

The remainder of this paper is organized as follows. After some motivations in this Introduction, Section 2, is devoted to preliminary results on $\phi$-divergence and the introduction of our estimator. Section 3 presents our new results on the Influence function. In Section 4, we investigate the finite-sample performance of the newly proposed estimators. To avoid interrupting the flow of the presentation, all technical arguments are deferred to the Appendix.

## 2. Extreme Value Statistics and $\phi$-Divergence Setting

A widely usen family of divergences is the so-called "*power divergences*", introduced by Cressie and Read [14] (see also Liese and Vajda [15, Chapter 2] and also the Renyi [16]'s paper is to be mentioned here), which are defined through the class of convex real-valued functions, for $\beta$ in $\mathbb{R} \setminus \{0, 1\}$:

$$x \in \mathbb{R}_+^* \longrightarrow \phi_\beta(x) := \frac{x^\beta - \beta x + \beta - 1}{\beta \cdot (\beta - 1)}. \tag{2.1}$$

We have $\phi_0(x) := -\log x + x - 1$ and $\phi_1(x) := x \log x - x + 1$. (For all $\beta \in \mathbb{R}$, we define $\phi_\beta(0) := \lim_{x \downarrow 0} \phi_\beta(x)$.) So, the $KL$-divergence is associated to $\phi_1$, the $KL_m$ to $\phi_0$, the $\chi^2$ to $\phi_2$, the $\chi_m^2$ to $\phi_{-1}$ and the Hellinger distance to $\phi_{1/2}$. In the monograph by Liese and Vajda [15] the reader may find detailed ingredients of the modeling theory as well as surveys of

the commonly used divergences. We recall some basic definitions for the readers' convenience. Unless otherwise specified we will assume that the function $\phi$ is a function of class $\mathcal{C}^2$, strictly convex, such that, for fixed $\widetilde{\boldsymbol{\theta}}$,

$$\int \left| \phi'\left(\frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}}(x)}\right) \right| d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x) < \infty, \quad \forall \boldsymbol{\theta} \in \Theta. \tag{2.2}$$

According to Broniatowski and Keziou [17], if the function $\phi(\cdot)$ satisfies the following conditions:

$$\text{there exists } 0 < \delta < 1 \text{ such that } \forall c \text{ in } [1 - \delta, \ 1 + \delta],$$

$$\text{we can find numbers } c_1, c_2, c_3 \text{ such that} \tag{2.3}$$

$$\phi(cx) \le c_1 \phi(x) + c_2 |x| + c_3, \quad \forall \text{ real } x,$$

then the assumption (2.2) is satisfied whenever $D_\phi(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) < \infty$, where $D_\phi(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ stands for the $\phi$-divergence between $\mathbb{P}_{\boldsymbol{\theta}}$ and $\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}$, we refer the reader to Broniatowski and Keziou [18, Lemma 3.2]. Also the real convex functions $\phi(\cdot)$ (2.1), associated with the class of power divergences, all satisfy the condition (2.2), including all standard divergences. Under assumption (2.2), using Fenchel duality technique, the divergence $D_\phi(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ can be represented as resulting from an optimization procedure,

$$\begin{aligned} D_\phi\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_0\right) &:= \int \phi\left(\frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}_0}(x)}\right) d\mathbb{P}_{\boldsymbol{\theta}_0}(x) \\ &= \sup_{\boldsymbol{\theta} \in \mathcal{U}} \int h\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}, x\right) d\mathbb{P}_{\boldsymbol{\theta}_0}(x), \end{aligned} \tag{2.4}$$

where $h(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}, \cdot) : x \mapsto h(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}, x)$ and

$$\begin{aligned} h\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}, x\right) &:= \int \phi'\left(\frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}}(x)}\right) d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x) dx \\ &\quad - \left[ \frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}}(x)} \phi'\left(\frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}}(x)}\right) - \phi\left(\frac{d\mathbb{P}_{\widetilde{\boldsymbol{\theta}}}(x)}{d\mathbb{P}_{\boldsymbol{\theta}}(x)}\right) \right]. \end{aligned} \tag{2.5}$$

This result was elegantly proven in, Keziou [19], Liese and Vajda [20] and Broniatowski and Keziou [17]. Broniatowski and Keziou [18] called it the dual form of a divergence, due to its connection with convex analysis. Furthermore, the supremum in this display (2.4) is unique and reached in $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, independently upon the value of $\widetilde{\boldsymbol{\theta}}$. Let $X_1, \ldots, X_n$ be an independent, identically distributed (i.i.d.) sample from an unknown distribution function (d.f.) $\mathbb{P}_{\boldsymbol{\theta}_0}$. Naturally, a class of estimators of $\widetilde{\boldsymbol{\theta}}_0$, called "dual $\phi$-divergence estimators" (D$\phi$DE's), is defined by

$$\widehat{\boldsymbol{\theta}}_\phi\left(\widetilde{\boldsymbol{\theta}}\right) := \arg\sup_{\boldsymbol{\theta} \in \Theta} \frac{1}{n} \sum_{i=1}^{n} h\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}, X_i\right), \quad \widetilde{\boldsymbol{\theta}} \in \Theta, \tag{2.6}$$

where $h(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is the function defined in (2.5). The class of estimators $\widehat{\boldsymbol{\theta}}_{\phi}(\widetilde{\boldsymbol{\theta}})$ satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} h\left(\widetilde{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}_{\phi}(\widetilde{\boldsymbol{\theta}}), X_i\right) = 0. \tag{2.7}$$

Formula (2.6) defines a family of $M$-estimators indexed by the function $\phi$ specifying the divergence and by some instrumental value of the parameter $\widetilde{\boldsymbol{\theta}}$. Application of dual representation of $\phi$-divergences has been considered by many authors; we cite among others, Keziou and Leoni-Aubin [21] for semiparametric two-sample density ratio models, bootstrapped $\phi$-divergences estimates are considered in Bouzebda and Cherfi [22], extension of dual $\phi$-divergences estimators to right censored data are introduced in Cherfi [23], for estimation and tests in copula models we refer to Bouzebda and Keziou [24, 25], and the references therein. Performances of dual $\phi$-divergence estimators for normal models are studied in Cherfi [26].

In what follows, we describe the procedure used to obtain the D$\phi$DE for the tail index. Let $X_1, \ldots, X_n$ be an independent, identically distributed (i.i.d.) sample from an unknown distribution function (d.f.) $\mathbb{P}$. Since it is well known that a distribution is in the domain of attraction of a Fréchet distribution if and only if the distribution has a regularly varying tail, we can assume that $1 - \mathbb{P}$ is regularly varying at $\infty$ with the exponent $-1/\gamma$, $\gamma$ is called the tail index of distribution $\mathbb{P}$:

$$\lim_{x \to \infty} \frac{1 - \mathbb{P}(ux)}{1 - \mathbb{P}(x)} = u^{-1/\gamma}, \quad \text{for every } u > 0, \tag{2.8}$$

or equivalently,

$$1 - \mathbb{P}(x) = x^{-1/\gamma} \ell(x), \tag{2.9}$$

where $\ell(x)$ is slowly varying at $\infty$, namely,

$$\lim_{x \to \infty} \frac{\ell(ux)}{\ell(x)} = 1, \quad \text{for every } u > 0. \tag{2.10}$$

In the Pareto-type case, the conditional distribution $\mathbb{P}_{Y_u}$ of relative excesses

$$Y_u = \left(\frac{X}{u} \mid X > u\right). \tag{2.11}$$

over a threshold $u$ satisfies

$$1 - \mathbb{P}_{Y_u}(y) = P\left(\frac{X}{u} > y \mid X > u\right) = y^{-1/\gamma} \frac{\ell(ux)}{\ell(x)}, \tag{2.12}$$

it is easily seen that ultimately

$$\mathbb{P}_{Y_u}(y) \longrightarrow 1 - y^{-1/\gamma} \quad \text{as } u \longrightarrow \infty. \tag{2.13}$$

A popular choice for the threshold $u$ in threshold based methods is $X_{n-k:n}$, the $k$th largest observation of the sample, with $k = k_n = \lfloor n\alpha \rfloor$ for some $\alpha \in (0,1)$. Here and elsewhere, $\lfloor u \rfloor$ denotes the largest integer $\leq u$. The quantile function pertaining to $\mathbb{P}$, is defined, for $u \in (0,1)$, by

$$\mathbb{Q}(u) = \inf\{x : \mathbb{P}(x) \geq u\}. \tag{2.14}$$

The empirical quantile function is given, for each $n \geq 1$ and $u \in (0,1)$, by

$$\mathbb{Q}_n(u) = \inf\{x : \mathbb{P}_n(x) \geq u\}. \tag{2.15}$$

The threshold $X_{n-k:n}$ is easily seen to be equal to $\mathbb{Q}_n(1 - (k/n))$. The idea of constructing the D$\phi$DE for the tail index is to assume the above Pareto approximation to hold exactly as a model for the conditional distribution of the relative excesses

$$Y_{jk} = \frac{X_{n-j+1:n}}{u}, \quad j = 1, \dots, k \text{ above a high threshold } u = X_{n-k:n}. \tag{2.16}$$

That is we can fit a Pareto model to the relative excesses. In this framework, the estimation of $\gamma$ can be handled through dual divergences techniques, which provide a wide range of estimators, including the Hill estimator, they all can be compared with respect to robustness properties. Consider the Pareto density

$$\left\{ p_\gamma(x) := \frac{x^{-1/\gamma - 1}}{\gamma} : x > 1; \ \gamma \in \mathbb{R}_+^* \right\}. \tag{2.17}$$

Specializing (2.4) to this setting, elementary calculation, for $\beta$ in $\mathbb{R} \setminus \{0,1\}$, gives

$$\frac{1}{\beta - 1} \int \left( \frac{d\mathbb{P}_{\tilde{\gamma}}(x)}{d\mathbb{P}_\gamma(x)} \right)^{\beta - 1} d\mathbb{P}_{\tilde{\gamma}}(x) dx = \left( \frac{\gamma}{\tilde{\gamma}} \right)^{(\beta - 1)} \left( \frac{\gamma}{\gamma\beta(\beta - 1) - \tilde{\gamma}(\beta - 1)^2} \right). \tag{2.18}$$

Using this last equality, one finds

$$\hat{D}_\beta(\tilde{\gamma}, \tilde{\gamma}_0) = \sup_\gamma \left\{ \left( \frac{\gamma}{\tilde{\gamma}} \right)^{(\beta - 1)} \left( \frac{\gamma}{\gamma\beta(\beta - 1) - \tilde{\gamma}(\beta - 1)^2} \right) \right.$$
$$\left. - \frac{1}{\beta k} \sum_{j=1}^{k} \left( \frac{\gamma}{\tilde{\gamma}} \right)^\beta Y_{jk}^{\{-\beta(1/\tilde{\gamma} - 1/\gamma)\}} - \frac{1}{\beta(\beta - 1)} \right\}. \tag{2.19}$$

We now consider an interesting particular case of the previous setup, for $\beta = 0$, one obtains

$$\widehat{D}_{KL_m}(\widetilde{\gamma}, \widetilde{\gamma}_0) := \sup_{\gamma} \left\{ -\frac{1}{k} \sum_{j=1}^{k} \left\{ \log\left(\frac{\gamma}{\widetilde{\gamma}}\right) - \left(\frac{1}{\widetilde{\gamma}} - \frac{1}{\gamma}\right) \log(Y_{jk}) \right\} \right\}, \tag{2.20}$$

which leads to the famous Hill estimator [1], given by

$$\widehat{\gamma}_k = \frac{1}{k} \sum_{j=1}^{k} \log_+(Y_{jk}) := \frac{1}{k} \sum_{j=1}^{k} \log_+(X_{n-j+1:n}) - \log_+(X_{n-k:n}), \tag{2.21}$$

independently upon $\widetilde{\gamma}$, where $\log_+ x = \log(\max(x, 1))$. Mason [27] show that consistency of the Hill estimator if $k = k_n$ is a sequence of positive integers satisfying

$$1 \le k_n \le n - 1, \quad k_n \longrightarrow \infty, \quad n^{-1} k_n \longrightarrow 0 \quad \text{as } n \longrightarrow \infty. \tag{2.22}$$

Further investigations concerning the asymptotic distribution of the Hill estimator have been made by Hall [28], Csörgő and Mason [29], Haeusler and Teugels [30], Beirlant and Teugels [31], and Bouzebda [32] ISUP. This is shown, under certain additional regularity conditions, on $\mathbb{P}$ and on $k = k_n$ satisfying (2.22).

## 3. Influence Function

In this section we study the robustness properties of the proposed estimators theoretically. In particular, we derive their influence functions from which the asymptotic variance follows. The following definition is needed for the statement of our forthcoming result. Recall that the influence function of a functional $T$ at a distribution $\mathbb{P}$ describes the effect on the estimate $T$ of an infinitesimal contamination to $\mathbb{P}$ at the point $x$ and is given by

$$\text{IF}(x; T, \mathbb{P}) = \lim_{\epsilon \to 0} \frac{T(\mathbb{P}_{\epsilon,x}) - T(\mathbb{P})}{\epsilon}, \tag{3.1}$$

where

$$\mathbb{P}_{\epsilon,x} = (1 - \epsilon)\mathbb{P} + \epsilon \delta_x \tag{3.2}$$

and $\delta_x$ is the Dirac measure putting all its mass at $x$ and $0 \le \epsilon \le 1$. In the following, we will derive the influence function for the functional form of the newly proposed estimator in an analogous way as for the classical $M$-estimators [33]. General results on influence functions

of the dual $\phi$-divergence estimators can be found in Toma and Broniatowski [34]. We will use the following notations

$$S := -\int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial^2}{\partial \gamma^2} h\left(\widetilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), z\right) d\mathbb{P}(z),$$

$$\psi_\alpha(\widetilde{\gamma}, \gamma, x)$$

$$:= \int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial t} \left[\frac{\partial}{\partial \gamma} h(\widetilde{\gamma}, T_\alpha(\mathbb{P}), t, z)\right]\Bigg|_{t=\mathbb{P}^{-1}(1-\alpha)} d\mathbb{P}(z) \left(\frac{\mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}} - \alpha}{p\left(\mathbb{P}^{-1}(1-\alpha)\right)}\right) \tag{3.3}$$

$$+ \frac{\partial}{\partial \gamma} h\left(\widetilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), x\right) \mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}}$$

$$- \frac{\partial}{\partial \gamma} h\left(\widetilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), \mathbb{P}^{-1}(1-\alpha)\right) \left(\mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}} - \alpha\right),$$

where $\mathbb{1}_A$ stands for the indicator function of the event $A$. Our results are summarized in the following theorem; its proof is given in the next section.

**Proposition 3.1.** *The influence function of the functional $T_\alpha(\mathbb{P}_n)$ corresponding to an estimator $\widehat{\gamma}_\phi(\widetilde{\gamma})$ is given by*

$$IF(x; T_\alpha(\mathbb{P}), \mathbb{P}) = S^{-1} \psi_\alpha(\widetilde{\gamma}, \gamma, x). \tag{3.4}$$

To illustrate the behavior of the obtained influence functions we restrict ourselves to the strict Pareto case; simulation results for other heavy-tailed distribution are presented in the next section. Figure 1 plots the influence functions of our estimators for the Pareto distribution. Observe that, for the Hellinger distance ($\beta = 0.5$) and the $\chi^2$-divergence ($\beta = 2$), the influence for the tail index $\gamma$ becomes negligible for large outliers. The influence functions are bounded, making the associated functionals robust, in contrast with the Hill estimator ($\beta = 0$) and the other divergences.

## 4. Simulation

In order to illustrate the robustness of the proposed statistical method, its finite sample behavior is investigated, both at contaminated as well as uncontaminated data. For the tail index $\gamma$, a comparison is made between the well-known Hill [1] estimator and the newly proposed estimator.

We consider 1000 simulated samples without contamination, each containing $n = 200$ observations, from the two Pareto-type distributions.

(i) The Fréchet distribution given by $\exp(-x^{-1/\gamma})$.

(ii) The Burr distribution given by

$$\left(\frac{\xi}{\xi + x^\tau}\right)^\lambda, \quad \text{such that } \gamma = \frac{1}{\lambda\tau}. \tag{4.1}$$
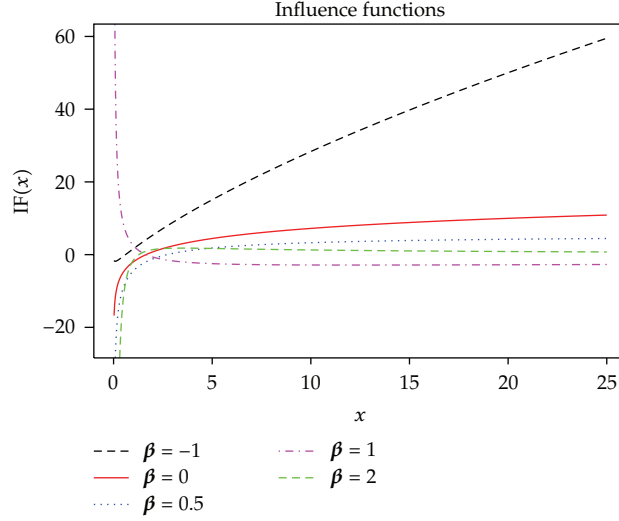
**Figure 1:** Influence functions for the Pareto distribution.

In the simulations, we have chosen $\gamma = 1/2$, for the Burr distribution $\xi = 10$, $\tau = 2$ $\lambda = 1$. The means of the D$\phi$DE (left) and the corresponding empirical mean squared errors (right), also as a function of $k$ are plotted. The horizontal line indicates the true value of $\gamma = 1/2$.

When the data are uncontaminated, although most robust estimators are known to be less efficient at the true model than maximum likelihood estimators [1], we notice that the $\gamma$ estimates seem to be fairly stable for intermediate values of $k$, making the influence of the choice of $k$ less troublesome and even with respect to mean squared error, the newly proposed estimator does not seem to lose too much accuracy, a close look to Figures 2 and 3 shows that there is a slight tendency to overestimation. The newly proposed D$\phi$DE's perform remarkably as well as the Hill [1] estimator.

However, a slight contamination (5%) is sufficient to make the D$\phi$DE associated to the $\chi^2$-divergence ($\beta = 2$) more appealing in terms of low MSE; see Figures 4, 5, 6, and 7. Furthermore when contamination increases, the D$\chi^2$DE performs remarkably better.

Overall, the simulation results in this section provide supporting evidence of the adequacy of the D$\phi$DE associated with the $\chi^2$-divergence with observations drawn from Fréchet and Burr distributions. Moreover, the sensitivity of this estimator for the choice of $k$ is low.

## Appendix

This section is devoted to the proofs of our result.

*Proof of Proposition 3.1.* For convenience, we recall the definition of the empirical measure $\mathbb{P}_n$ associated with the random variables $X_i$, $i = 1, \ldots, n$, which is given by

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{n-i+1:n}}. \tag{A.1}$$

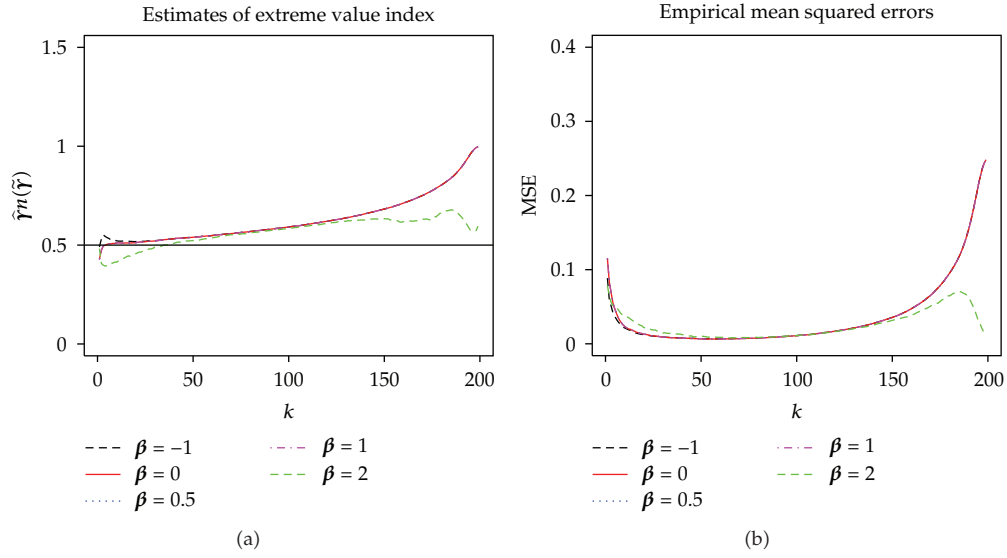**Figure 2:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Fréchet distribution.
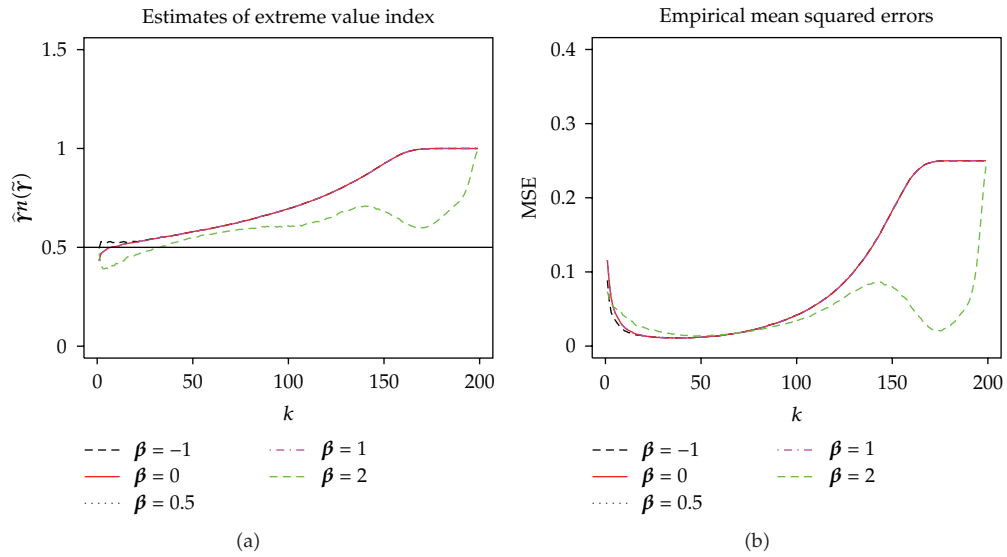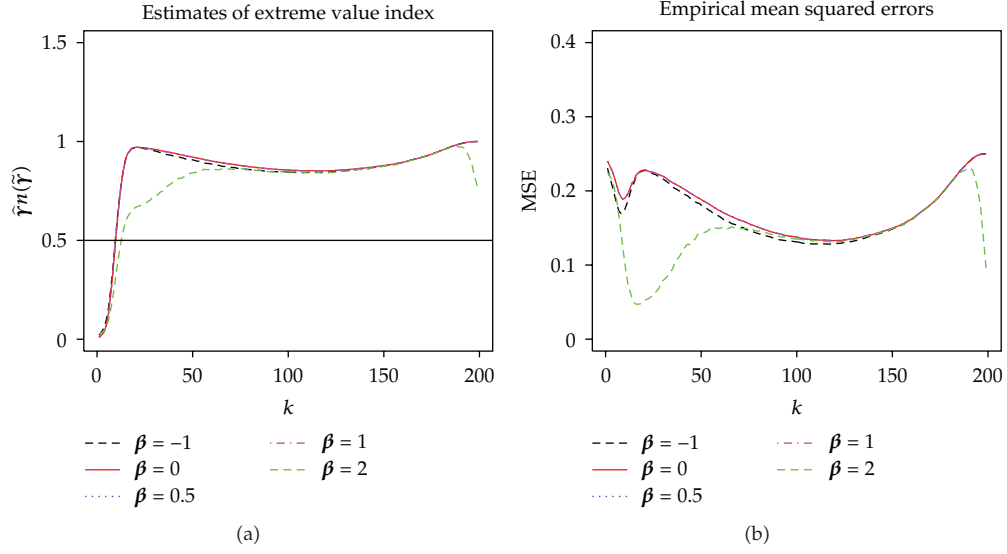


**Figure 3:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Burr distribution.

We define the estimator as the value $\widehat{\gamma}_\phi(\widetilde{\gamma})$ which maximizes, independently of $(k/n)$, the following estimating equation:

$$\int_{\mathbb{Q}_n(1-k/n)}^{\infty} h\left(\widetilde{\gamma}, \gamma, \mathbb{Q}_n\left(1-\frac{k}{n}\right), x\right) d\mathbb{P}_n(x), \qquad (\text{A.2})$$

**Figure 4:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Fréchet distribution, with $\epsilon = 0.05$.
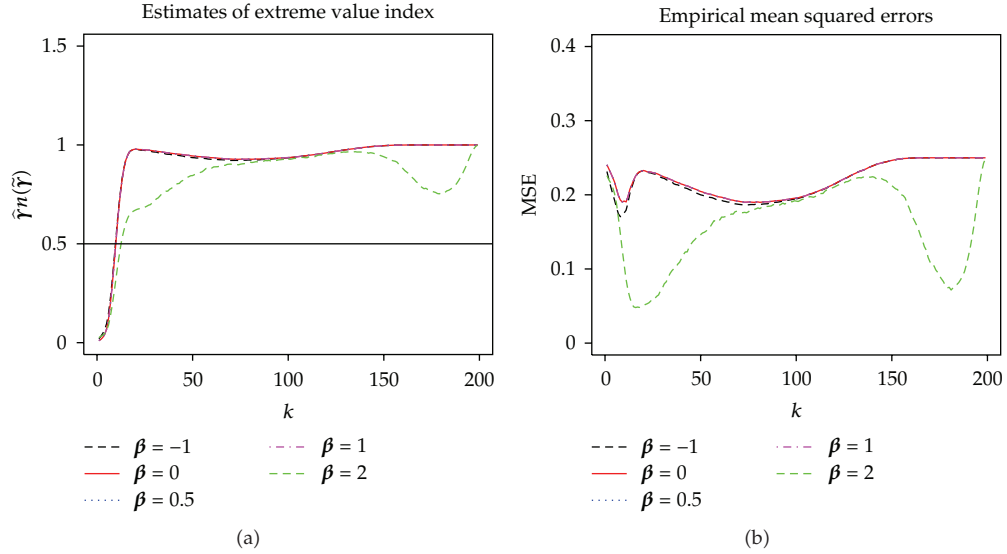


**Figure 5:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Burr distribution, with $\epsilon = 0.05$.

or, equivalently, as the solution, in $\gamma$, of the following equation:

$$\int_{\mathbb{Q}_n(1-k/n)}^{\infty} \frac{\partial}{\partial \gamma} h\left(\tilde{\gamma}, \gamma, \mathbb{Q}_n\left(1 - \frac{k}{n}\right), x\right) d\mathbb{P}_n(x) = 0. \qquad (A.3)$$
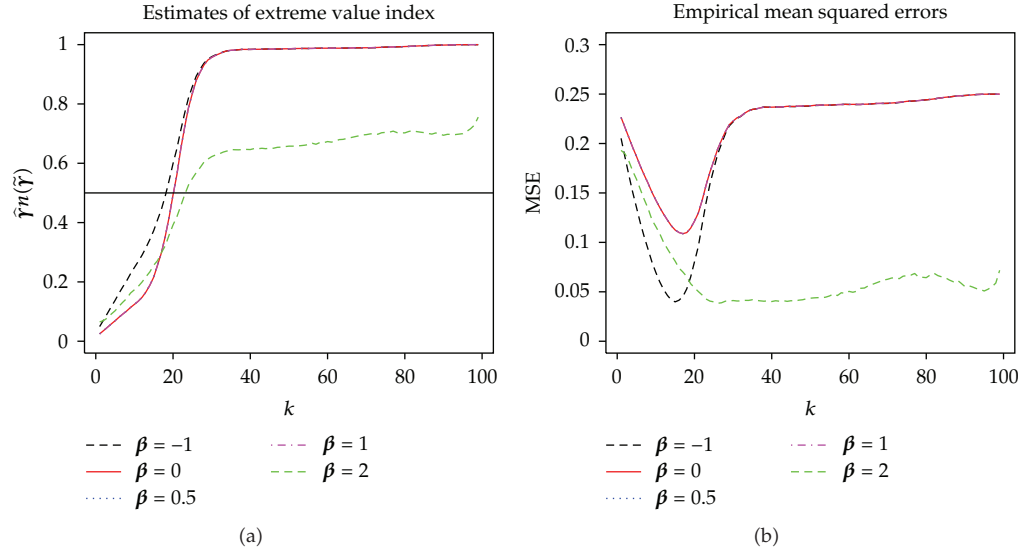
**Figure 6:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Fréchet distribution, with $\epsilon = 0.2$.
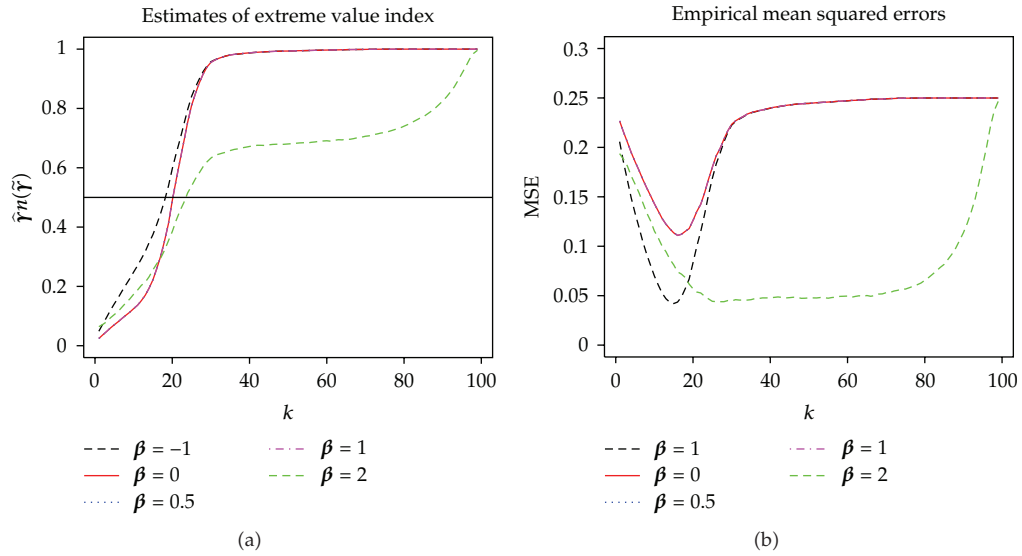


**Figure 7:** Mean $\gamma$ estimates, corresponding to empirical mean squared errors as a function of $k$ for the Burr distribution, with $\epsilon = 0.2$.

In this view, the estimator may be written in the form of a functional $T_\alpha(\mathbb{P}_n)$, given by

$$\int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\left(\tilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), x\right) d\mathbb{P}(x) = 0. \qquad (A.4)$$

We continue by rewriting (A.4) for the contaminated distribution as given in the definition of the influence function defined in (3.1), that is,

$$
(1 - \epsilon) \int_{\mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z\Big) d\mathbb{P}(z)
$$
$$
+ \epsilon \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), x\Big) \mathbb{1}_{\{x \geq \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)\}} = 0.
$$

(A.5)

Observe that

$$
\frac{\partial}{\partial \epsilon} \left[ \epsilon \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), x\Big) \mathbb{1}_{\{x \geq \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)\}} \right] \Bigg|_{\epsilon=0}
$$
$$
= \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), x\Big) \mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}},
$$

(A.6)

$$
\frac{\partial}{\partial \epsilon} \left[ (1 - \epsilon) \int_{\mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z\Big) d\mathbb{P}(z) \right] \Bigg|_{\epsilon=0}
$$
$$
= \frac{\partial}{\partial \epsilon} \left[ \int_{\mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z\Big) d\mathbb{P}(z) \right] \Bigg|_{\epsilon=0}
$$

(A.7)

$$
- \int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), z\Big) d\mathbb{P}(z).
$$

Keeping in mind the definition of $T_\alpha$ as in (A.4), the last term in (A.7) disappears. We next evaluate the first term in the right side of (A.7). We infer readily by using Leibnitz's integral rule, that:

$$
\frac{\partial}{\partial \epsilon} \left[ \int_{\mathbb{P}_{\epsilon,x}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z\Big) d\mathbb{P}(z) \right] \Bigg|_{\epsilon=0}
$$
$$
= \int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \epsilon} \left[ \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z\Big) \right] \Bigg|_{\epsilon=0} d\mathbb{P}(z)
$$

(A.8)

$$
- \frac{\partial}{\partial \gamma} h\Big(\tilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), \mathbb{P}^{-1}(1-\alpha)\Big) \Big( \mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}} - \alpha \Big).
$$

In a similar way, we can therefore write,

$$
\int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial \epsilon} \left[ \frac{\partial}{\partial \gamma} h\left( \widetilde{\gamma}, T_\alpha(\mathbb{P}_{\epsilon,x}), \mathbb{P}_{\epsilon,x}^{-1}(1-\alpha), z \right) \right] \Bigg|_{\epsilon=0} d\mathbb{P}(z)
$$

$$
= \left[ \int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial^2}{\partial \gamma^2} h\left( \widetilde{\gamma}, T_\alpha(\mathbb{P}), \mathbb{P}^{-1}(1-\alpha), z \right) d\mathbb{P}(z) \right] \frac{\partial}{\partial \epsilon} T_\alpha(\mathbb{P}_{\epsilon,x}) \Bigg|_{\epsilon=0} \tag{A.9}
$$

$$
+ \int_{\mathbb{P}^{-1}(1-\alpha)}^{\infty} \frac{\partial}{\partial t} \left[ \frac{\partial}{\partial \gamma} h(\widetilde{\gamma}, T_\alpha(\mathbb{P}), t, z) \right] \Bigg|_{t=\mathbb{P}^{-1}(1-\alpha)} d\mathbb{P}(z) \left( \frac{\mathbb{1}_{\{x \geq \mathbb{P}^{-1}(1-\alpha)\}} - \alpha}{p\left(\mathbb{P}^{-1}(1-\alpha)\right)} \right).
$$

The proof of Proposition 3.1 is therefore completed. $\qquad\square$

## Acknowledgments

## References

[1] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[2] R. L. Smith, "Estimating tails of probability distributions," *The Annals of Statistics*, vol. 15, no. 3, pp. 1174–1207, 1987.

[3] C. Alexander, "Assessment of operational risk capital. newblock," in *Riskman-Agement, Challenge and Opportunity*, M. Frenkel, U. Hommel, and M. Rudolf, Eds., 2nd edition, 2005.

[4] R. Dell'Aquila and P. Embrechts, "Extremes and robustness: a contradiction?" *Financial Markets and Portfolio Management*, vol. 20, no. 1, pp. 103–118, 2006.

[5] V. Brazauskas and R. Serfling, "Robust estimation of tail parameters for twoparameter and exponential models via generalized quantile statistics," *Extremes*, vol. 3, pp. 231–249, 2000.

[6] V. Brazauskas and R. Serfling, "Robust and effcient estimation of the tail index of a single-parameter Pareto distribution," *The North American Actuarial Journal*, vol. 4, pp. 12–27, 2000.

[7] D. J. Dupuis and C. A. Field, "Robust estimation of extremes," *The Canadian Journal of Statistics*, vol. 26, no. 2, pp. 199–215, 1998.

[8] L. Peng and A. H. Welsh, "Robust estimation of the generalized Pareto distribution," *Extremes*, vol. 4, no. 1, pp. 53–65, 2001.

[9] S. F. Juárez and W. R. Schucany, "Robust and efficient estimation for the generalized Pareto distribution," *Extremes*, vol. 7, no. 3, pp. 237–251, 2004.

[10] B. Vandewalle, J. Beirlant, A. Christmann, and M. Hubert, "A robust estimator for the tail index of Pareto-type distributions," *Computational Statistics & Data Analysis*, vol. 51, no. 12, pp. 6252–6268, 2007.

[11] M. Kim and S. Lee, "Estimation of a tail index based on minimum density power divergence," *Journal of Multivariate Analysis*, vol. 99, no. 10, pp. 2453–2471, 2008.

[12] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[13] M. Hubert, G. Dierckx, and D. Vanpaemel, "Detecting inuential data points for the Hill estimator in Pareto-type distributions," *Computational Statistics & Data Analysis*. In press.

[14] N. Cressie and T. R. C. Read, "Multinomial goodness-of-fit tests," *Journal of the Royal Statistical Society B*, vol. 46, no. 3, pp. 440–464, 1984.

[15] F. Liese and I. Vajda, *Convex Statistical Distances*, vol. 95 of *Teubner Texts in Mathematics*, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, Germany, 1987.

[16] A. Rényi, "On measures of entropy and information," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561, University of California Press, Berkeley, Calif, USA, 1961.

[17] M. Broniatowski and A. Keziou, "Parametric estimation and tests through divergences and the duality technique," *Journal of Multivariate Analysis*, vol. 100, no. 1, pp. 16–36, 2009.

[18] M. Broniatowski and A. Keziou, "Minimization of $\varphi$-divergences on sets of signed measures," *Studia Scientiarum Mathematicarum Hungarica*, vol. 43, no. 4, pp. 403–442, 2006.

[19] A. Keziou, "Dual representation of $\varphi$-divergences and applications," *Comptes Rendus Mathématique*, vol. 336, no. 10, pp. 857–862, 2003.

[20] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, 2006.

[21] A. Keziou and S. Leoni-Aubin, "On empirical likelihood for semiparametric two-sample density ratio models," *Journal of Statistical Planning and Inference*, vol. 138, no. 4, pp. 915–928, 2008.

[22] S. Bouzebda and M. Cherfi, "General bootstrap for dual $\varphi$-divergence estimates," *Journal of Probability and Statistics*, vol. 2012, Article ID 834107, 33 pages, 2012.

[23] M. Cherfi, "Dual divergences estimation for censored survival data," *Journal of Statistical Planning and Inference*, vol. 142, no. 7, pp. 1746–1756, 2012.

[24] S. Bouzebda and A. Keziou, "New estimates and tests of independence in semiparametric copula models," *Kybernetika*, vol. 46, no. 1, pp. 178–201, 2010.

[25] S. Bouzebda and A. Keziou, "A new test procedure of independence in copula models via $\chi^2$-divergence," *Communications in Statistics*, vol. 39, no. 1-2, pp. 1–20, 2010.

[26] M. Cherfi, "Dual $\phi$-divergences estimation in normal models," *Computation*. In press. http://arxiv.org/abs/1108.2999.

[27] D. M. Mason, "Laws of large numbers for sums of extreme values," *The Annals of Probability*, vol. 10, no. 3, pp. 754–764, 1982.

[28] P. Hall, "On some simple estimates of an exponent of regular variation," *Journal of the Royal Statistical Society B*, vol. 44, no. 1, pp. 37–42, 1982.

[29] S. Csörgő and D. M. Mason, "Central limit theorems for sums of extreme values," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 98, no. 3, pp. 547–558, 1985.

[30] E. Haeusler and J. L. Teugels, "On asymptotic normality of Hill's estimator for the exponent of regular variation," *The Annals of Statistics*, vol. 13, no. 2, pp. 743–756, 1985.

[31] J. Beirlant and J. L. Teugels, "Asymptotics of Hill's estimator," *Teoriya Veroyatnosteǐ i ee Primeneniya*, vol. 31, no. 3, pp. 530–536, 1986.

[32] S. Bouzebda, "Bootstrap de l'estimateur de Hill: théorèmes limites," *Annales de l'I.S.U.P.*, vol. 54, no. 1-2, pp. 61–72, 2010.

[33] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1986.

[34] A. Toma and M. Broniatowski, "Dual divergence estimators and tests: robustness results," *Journal of Multivariate Analysis*, vol. 102, no. 1, pp. 20–36, 2011.